charles river analytics

Charles River Analytics Inc.
625 Mount Auburn St.
Cambridge, MA 02138
www.cra.com

p: 617.491.3474
f: 617.868.0780

20 February 2015

Dr. Harold Hawkins
ONR Code 341
Office of Naval Research
875 North Randolph SL
Arlington. VA 22203-1995

Reference: US Navy Contract N00014-12-C-0653: "The Model Analyst's Toolkit: Scientific Model Development, Analysis, and Validation"
Charles River Analytics Contract No. C12186

Subject: Contractor's Quarterly Status Report #10
Reporting Period: 20-November-2014 to 19-February-2015

Dear Dr. Hawkins,

Please find enclosed 1 copy of the Quarterly Status Report for the referenced contract. Please feel free to contact me with any questions regarding this report or the status of the "The Model Analyst's Toolkit: Scientific Model Development, Analysis, and Validation" effort.

Sincerely,

W. Scott Neal Reilly

W. Scott Neal Reilly
Principal Investigator

cc: Cheryl Gonzales, DCMA
Annetta Burger, ONR
Whitney McCoy, Charles River Analytics

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **20 FEB 2015**	2. REPORT TYPE	3. DATES COVERED **20-11-2014 to 19-02-2015**

4. TITLE AND SUBTITLE **The Model Analyst's Toolkit: Scientific Model Development, Analysis, and Validation**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Charles River Analytics Inc.,,625 Mount Auburn St.,,Cambridge,,MA,02138**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**13**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Charles River Analytics

Monthly Technical Progress Report No. R12186-10
Reporting Period: November 20, 2014 to February 19, 2015
Government Contract No. N00014-12-C-0653
Charles River Analytics Contract No. C12186

The Model Analyst's Toolkit: Scientific Model Development, Analysis, and Validation

Quarterly Status Report

Principal Investigator: Scott Neal Reilly

Charles River Analytics
625 Mount Auburn Street
Cambridge, MA 02138
617-491-3474

February 20, 2015

1. Executive Summary

The proposed research effort builds on and extends the work of the previous ONR-funded "Validation Coverage Toolkit for HSCB Models" project. The overall objectives of the on-going research program are:

- Help scientists create, analyze, refine, and validate rich scientific models
- Help computational scientists verify the correctness of their implementations of those models
- Help users of scientific models, including decision makers within the US Navy, to use those models correctly and with confidence
- Use a combination of human-driven data visualization and analysis, automated data analysis, and machine learning to leverage human expertise in model building with automated analyses of complex models against large datasets

Specific objectives for the current effort include:

- **Fluid temporal correlation analysis.** Our objective is to design a new method for performing temporally fluid correlation analysis for temporal sets of data and implement the method as a new prototype component within the Model Analyst's Toolkit (MAT) software application.
- **Automated suggestions for model construction and refinement.** Our objective is to design and implement a prototype mechanism that learns from data how factors interact in non-trivial ways in scientific models.
- **Data validation and repair.** Our objective is to design and implement a prototype capability to identify likely errors in data based on anomalies relative to historic data and to use models of historic data to offer suggested repairs.
- **System prototyping.** Our objective is to incorporate all improvements into the MAT software application and make the resulting application available to the government and academic research community for use in scientific modeling projects.
- **Evaluation of applicability to multiple scientific domains.** Our objective is to ensure (and demonstrate) that MAT can be applied to a wide range of scientific domains by identifying and building at least one neurological and/or physiological model and analyze the associated data with MAT, making any extensions to the MAT tool that are needed to support the analysis of such a model.

2. Overview of Problem and Technical Approach

2.1. Summary of the Problem

One of the most powerful things scientists can do is to create models that describe the world around us. Models help scientists organize their theories and suggest additional experiments to run. Validated models also help others in more practical applications. For instance, in the hands of military decision makers, human social cultural behavior (HSCB) models can help predict instability and the socio-political effects of missions, whereas models of the human brain and mind can help educators and trainers create curricula that more effectively improve the knowledge, skills, and abilities of their pupils.

While there are various software tools that are used by the scientific community to help them develop and analyze their models (e.g., Excel, R, Simulink, Matlab), they are largely so general in purpose (e.g., Excel, R) or so focused on computational models in particular (e.g., Simulink, Matlab), that they are not ideal for rapid model exploration or for use by non-computational scientists. They also largely ignore the problem of validating the models, especially when the models are positing causal claims as most interesting scientific models do. To address this gap, Charles River Analytics undertook the "Validation Coverage Toolkit for HSCB Models" project with ONR. Under this effort, we successfully designed, implemented, informally evaluated, and deployed a tool called the Model Analyst's Toolkit (MAT), which focused on supporting social scientists to visualize and explore data, develop causal models, and validate those models against available data (Neal Reilly, 2010; Neal Reilly, Pfeffer, Barnett et al., 2011, 2010).

As part of the development of the MAT tool, we identified four important extensions to that research program that would further support the scientific modeling process:

- Correlation analyses are still the standard way of identifying relationships between factors in a model, but correlations are fundamentally flawed as a tool for analyzing potentially causal or predictive relationships as they assume instantaneous effects. Even performing correlation analyses with a temporal offsets between streams of data is insufficient as the temporal gap between the causal or predictive event and the following event may not be the same every time (either because of variability in the system being modeled or because of variability introduced by a fixed sampling rate). What we need is a novel way of evaluating the true predictive power across streams of data that can deal with fluid offsets between changes in one stream of data and follow events in the other stream of data.
- Modeling complex phenomena is a fundamentally difficult task. Human intuition and analysis is by far the most effective way of performing this task, but even humans can be overwhelmed by the complexity of modeling the systems they are studying (e.g., socio-political system, human neurophysiology). Automated tools, while not especially good at generating reasonable scientific hypotheses, *are* extremely good at processing large amounts of data. We believe there is an opportunity for computational systems to enhance human scientific inquiry. Under the "Validation Coverage Toolkit for HSCB Models" project, we demonstrated how automated tools could help human scientists to analyze and validate their models against data. We believe a similar approach can be used to help suggest modifications to the human-built models to make them better match the available data. To be useful, however, such automated analyses will need to be rich enough to suggest subtle data interactions that are most likely to be missed by the human scientist. For instance, correlations (especially correlations that take into account fluid temporal displacements) could be used to identify likely relationships between streams of data, but such an approach would miss complex, non-linear relationships between interrelated factors that cannot be effectively analyzed with

simple two-way correlations. For instance, if crime waves are associated with increases in unemployment *or* drops in the police presence, that would be hard to identify with a correlation analysis. We need richer automated data analysis techniques that can extract complex, non-linear, multi-variable relationships between data if we are to effectively suggest model improvements to human scientists.

- Even if a scientific model is sound, if the data sets provided as inputs to the model are unreliable, the results of the model are still suspect. And, unfortunately, data will often be wrong. For instance, HSCB surveys are notoriously unreliable and biased for a variety of reasons, and neurological and physiological data can be corrupted by broken or improperly used sensors. If it were possible to identify when data was unreliable and, ideally, even repair the data, then the models that are using the data could once again be effectively used.
- The MAT tool we developed under the "Validation Coverage Toolkit for HSCB Models" project was focused primarily on assisting social scientists in the analysis, refinement, and validation of HSCB models. In parallel with that effort, however, we also took an opportunity to apply MAT to evaluating neurological and physiological data under the DARPA-funded CRANIUM (Cognitive Readiness Agents for Neural Imaging and Understanding Models) program. We discovered the generality of the MAT tool makes it potentially applicable to a great number of different scientific domains. MAT proved to be a useful, but peripheral tool, in CRANIUM. We believe MAT could be applied to a broader suite of scientific modeling problems than it has been so far.

2.2. Summary of our Approach

To address these identified gaps and opportunities, we are extending MAT's support for model development, analysis, refinement, and validation; enhancing MAT to analyze and repair data; and demonstrating MATs usefulness in additional scientific modeling domains. Our approach encompasses the following four areas, which correspond to the four gaps/opportunities identified in the previous section:

- **Temporally Fluid Correlation Analysis.** We are designing a new method to perform Temporally Fluid Correlational Analysis on temporal sets of data, and we are implementing the method as a new component within the MAT software application. The version of MAT at the beginning of the new effort supported correlation analysis for temporally offset data; it shifts the two data streams being compared by a fixed offset that is based on the sampling rate of the data (i.e., data that is sampled annually will be shifted by one year at a time), performs a standard correlation on the shifted data, plots the correlation value against the amount of the offset, and then repeats the process for the next offset amount. If two data streams are shifted by a fixed offset (e.g., changes in one stream are always followed by a comparable value in the other stream after a fixed time), then this method will find that offset. Under the current effort, we are expanding on this capability to support fluid temporal shifts within the data streams. That is, we are making it possible to identify when the temporal offset between the change in the first data stream and its effect in the second stream is not a static amount of time.

- **Automated suggestions for model construction and refinement.** We are designing and implementing a mechanism to learn how factors interact in non-trivial ways in scientific models. In particular, we are developing a method for learning disjuncts, conjuncts, and negations. This mechanism starts with the model developed by the scientist user and make recommendations for possible adjustments to make it more complete by performing statistical data mining and machine learning.
- **Data validation and repair.** Recognizing that data contains errors is plausible once we understand the relationships between data sets. That is, if we are able to develop models of the correlations between sets of data, then we can build systems that notice when these correlations do not hold in new data, indicating possible errors in data. For instance, if we know that public sentiment tends to vary similarly between nearby towns, then when one town shows anomalous behavior, we can reasonably suspect problems with the data. There might be local issues that cause the anomaly, but it is, at least, worth noting and bringing to the attention of the user of the data and model. As MAT is designed to help analyze models and recognize inter-data relationships, it is primed to perform exactly this analysis. Existing methods perform similar types of analysis for environmental data (Dereszynski & Dietterich, 2007, 2011). For instance, a broken thermometer can be identified and the data from it even estimated by looking at the temperature readings of nearby thermometers, which will generally be highly correlated.
- **Application to multiple scientific modeling domains.** To ensure (and demonstrate) that MAT can be applied to a wide range of scientific domains, we are identifying and building at least one neurological and/or physiological model and analyzing the associated data with MAT, making any extensions to the MAT tool that are needed to support the analysis of such a model. The initial MAT effort focused on HSCB models; by focusing this effort on harder-science models at much shorter time durations, we believe we can effectively evaluate an interesting range of applications of the MAT tool.

3. Current Activities and Status

During the current reporting period, we focused primarily on improving the causal model validation capabilities in MAT. Our basic approach is to provide a toolbox of methods that can help to analyze and invalidate different types of causal claims in different situations.

We have previously implemented and integrated analyses based on: Pearson correlation, static offset correlation, and dynamic offset correlation (based on gait recognition techniques). We have also implemented analysis based on Granger causality, though we have not previously integrated it into the MAT front end. We describe our efforts to perform this integration during this period in Section 3.1. We also created a Matlab implementation of Convergent Cross Mapping which we plan to port and integrate during the next period.

We also explored new techniques that are being developed by the UAI community that are finding novel ways of extracting causal information from even observational data. Some of these are not appropriate for MAT (e.g., they work on relational instead of temporal data) and others we are not likely to be able to implement with the current budget and time, but we are starting to explore follow-on opportunities with, among others, ONR, DARPA, and IARPA, that we hope might support implementing some of these concepts as part of MAT or as part of similar analysis tool.

Another focus of our effort this period was on simplifying the workflow for using multiple datasets, such as a development dataset and a testing dataset. This effort is described in Section 3.2.

3.1. Improvements in Granger Causality User Interface

Various metrics of causality are being integrated within MAT, including Granger causality. Granger causality tests whether a data series helps when predicting future values of another data series. More specifically, if the prediction of a data series is improved in a statistically significant way using a second data series, then the second data series is considered a Granger-cause of the first. However, whether there is a Granger-cause relationship between two data series is also dependent on the temporal lag used for the calculations. Therefore, MAT will allow the user to adjust the temporal lag and see what effect it has on the Granger causality result. This will allow the user to make the determination of whether the temporal lag is of appropriate scale for there to be a potential causal relationship. This kind of user interface is similar to the one we use for interacting with the dynamic offset correlation and we are building off of that in our new feature as well. We have an initial implementation that we are refining and should be completed in the next period.

3.2. Improvements in Causal Model Validation

One key feature of MAT is the ability to validate scientific models against various data sets. This is important when, as is common, there are different data sets for creating the model, sometimes another for refining it, and a separate one for validation. This means that we need to make it as easy as possible to specify which data set to use for analysis and to support multiple data sets at once for maximal ease of use. That is, allowing the user to choose subsets of their data to be used during validation is important when testing the validity of a causal model in multiple contexts. For example, a causal model that studies the causes of an increase in crime may want to be validated across all countries or within a single country. Therefore, the user interface is being updated so the user can choose what data categories are used during validation. Additionally, new categories will be able to be created so that the user can reorganize their data and use these new categories as the input to validation. The user will also be able to compare the validation of two (or more) subsets of their data at the same time. This will allow for side-by-side comparison of the validity of the causal model using different sets of data.

4. Planned Activities

During the upcoming reporting period, we plan to focus on the following tasks:

- Completing work on the causal analysis implementation in MAT including completing the integration of Granger causality, porting Convergent Cross Mapping from Matlab to java and integrating it, and (if time allows) providing a reporting mechanism that combines the results of the various causal analysis mechanisms in a single place.
- Beginning work on data validation. We had hoped to begin work on this during the current period, but the causal analysis work took more time than anticipated due to some bugs in a third-party library that we found and fixed, but that slowed our overall progress.

5. Evaluation and Transition

We continue to focus on making MAT available to the government and academic research communities and to look for opportunities to use MAT on a variety of ongoing research efforts.

Our abstract on “Tools for Validating Causal and Predictive Claims in Social Science Models” was accepted for presentation at the 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) in July, so we have begun work on the full paper submission which is due during the next reporting period.

Also, with the new release of MAT, we reached out to people who have previously requested copies of the MAT software. We heard back from Hasan Davulcu at Arizona State and Rick Grannis at UCLA, both of whom have expressed interest in trying out the latest release of MAT. Erin Fitzgerald, the Program Director of the Minerva Research Initiative, has also agreed to include an announcement about the new release of MAT in her next program-wide announcements email.

In an effort to continue to build awareness of the effort, we also announced the new release through the Charles River Facebook page, purchased a small advertising campaign aimed at data scientists on Facebook, and produced a press release that was sent to a number of news outlets (it was picked up by an online-news wire service), and we placed blurb/link to the press release on our corporate website home page.

We have also used the explorations into causal analysis and validation done under MAT as the basis for seedling pitches to DARPA (Steve Jameson) and IARPA (Steve Rieber), both of whom we have spoken to and have expressed initial interest in the MAT work and pursuing follow-on ideas. We have not included these opportunities yet in the table below as they are still fairly recent and uncertain. We will add them to the table in the next report if they proceed into more significant prospects.

Table 1 summarizes our transition progress to date. We will continue to update this table as we make additional progress and will include it as a regular part of future status reports.

Program	Customer	Comments
On-going efforts		
Tourniquet Master Trainer (TMT) (Phase II SBIR)	US Army's Telemedicine & Advanced Technology Research Center (TATRC)	MAT is being used to visualize and analyze data from sensors on a medical manikin that indicate whether a number of novel medical devices used to combat junctional and inguinal hemorrhaging are being applied properly. This is an ongoing program.
Laparoscopic Surgery Training System (LASTS) (Phase II SBIR)	US Navy's Office of Naval Research (ONR)	Under lasts, Charles River and Caroline Cao at Wright State University are using MAT to analyze data collected from the location of the laproscopic surgery tools tools during an experiment. Surgical tools are instrumented with markers and 3D data is collected on their location as the person performs the task. This is a now-completed program.
Cognitive Readiness Agents for Neural Imaging and Understanding Models (CRANIUM) (Phase I SBIR)	US Navy's Office of Naval Research (ONR)	MAT was used to visualize and extract patterns of stress and workload from neuro-physiological data for training systems. This was a Phase I SBIR program that did not progress to Phase II.
Business Intelligence Visualization for Organizational Understanding, Analysis, and Collaboration (BIVOUAC) Phase II SBIR	US Navy's Space and Naval Warfare Systems Command (SPAWAR)	MAT is being evaluated as part of the BIVOUAC SBIR program, which provides data analysis and visualization for Enterprise Resource Planning (ERP) systems for the Navy. This is an ongoing Phase II SBIR program.

Adaptive toolkit for the Assessment and augmentation of Performance by Teams in Real time (ADAPTER) (Phase I SBIR)	US Air Force Research Lab Human Effectiveness Directorate (AFRL/RH)	MAT is being used to analyze neuro-physiological data from cyber operators to evaluate cognitive workload during team-based cyber operations. This is an ongoing Phase II SBIR program.
Anticipated Efforts		
Enhancing Intuitive Decision Making Through Implicit Learning (I2BRC) (ONR Basic Research Challenge BAA)	US Navy's Office of Naval Research (ONR) Charles River is a subcontractor to DSCI MESH Solutions, LLC	The intention is to use MAT to help analyze neuro-physiological data to help better understand how implicit learning and intuitive decision making work. This is an ongoing BAA program. We recently received our first data to review, though the first batch did not include temporal data that could leverage MAT. This effort is going on temporary hiatus as ONR changes the structure of the program.
A system for augmenting training by Monitoring, Extracting, and Decoding Indicators of Cognitive Load (MEDIC)	US Army's Telemedicine & Advanced Technology Research Center (TATRC)	We are evaluating the practicability of using MAT to analyze and visualize neuro-physiological data from combat medic trainees to identify periods of stress and cognitive overload. This is an active Phase II SBIR program where MAT is being considered for data analysis.

Table 1. MAT Transition and Use Progress

In addition we have provided copies of MAT to the following institutions based on their requests for the software: the University of Michigan, Arizona State University, Kansas State University, University of California at Los Angeles, the Naval Medical Research Unit at Wright Patterson Air Force Base, Concordia University (Montreal), the University of Wisconsin, the University of Maryland, and the Air Force Research Laboratory's Human Effectiveness Directorate, the Intelligence Advanced Research Projects Agency (IARPA), and the Joint Advanced Warfighting Division (JAWD).

6. Budget and Project Tracking

As of February 28, 2015, we have spent $751,265, or 81% of our total budget of $928,224, in 83% of the scheduled time. Our current funding is $862,477, so we have spent 87% of our available funding. Note that these numbers include the 26-NOV-2014 funding increment.

Overall, we believe we are in good shape to complete the project on time and on budget.

7. References

Dereszynski, E. W.& Dietterich, T. G. (2011). Spatiotemporal models for data-anomaly detection in dynamic environmental monitoring campaigns. *ACM Transactions on Sensor Networks (TOSN)*, 8.

Dereszynski, E.& Dietterich, T. (2007). Probabilistic models for anomaly detection in remote sensor data streams.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424-438.

Granger, C. W. (1980). Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and Control*, 2, 329-352.

Manning, C., Raghavan, R., and Schutze, H. (2008). *Introduction to Information Retrieval*.: Cambridge University Press.

Myers, C.& Rabiner, L. (1981). A level building dynamic time warping algorithm for connected word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29, 284-297.

Neal Reilly, S. (2010). *Validation Coverage Toolkit for HSCB Models*. Charles River Analytics Inc.

Neal Reilly, W. S., Pfeffer, A., and Barnett, J. (2010). A Metamodel Description Language for HSCB Modeling. In *Advances in Cross-Cultural Decision Making*.

Neal Reilly, W. S., Pfeffer, A., Barnett, J., Chamberlain, J., and Casstevens, R. (2011). A Computational Toolset for Socio-Cultural Data Exploration, Model Refinement, and Model Validation. In *Human Social Culture Behavior (HSCB) Focus*.

Olszewski, R. (2001). *Generalized feature extraction for structural pattern recognition in time-series data*. Carnegie Mellon University, School of Computer Science.

Salvador, S.& Chan, P. (2007). Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11, 561-580.

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., and Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, 338, 496-500.